Open source tools for content management

Jesús Tramullas

Citación recomendada: Jesús Tramullas. *Open source tools for content management* [en linea]. "Hipertext.net", num. 3, 2005. <http://www.hipertext.net> [Consulted: 12 feb. 2007]. .

1. Information Management and Content Management

Information and documentation services available on the Internet through web servers are growing in an exponential manner. The logical evolution of the Internet over the last 10 years has been producing a replacement of static web pages and documents by dynamically generated documents. This is due both to user interaction with work processes and flows defined by service creators and to the availability of growing information repositories. This has meant a progressive evolution from a concept of web page publishing which was quite simple in its origins to more complex and differentiated schemes relying on procedures and techniques based on information management. The increasing complexity of services and systems supporting them has made it necessary to formulate a theoretical and practical corpus capable of combining classical information management techniques within organizations with the particular features of the digital environment.

This evolution which has been accelerating during the first years of the 20 th Century has had an impact not only information management methods and techniques but also on the very technology used for information management; and consequently the market for products and services (GILBANE, 2000). If it is true that during the second half of the 1990's we could distinguish between products for document management, information retrieval, etc., since the turn of the century there has been a convergence between all platforms. Nowadays it has become easy to find solutions that intend to be global and to provide support for the whole process of information management within an organization. The kind of tools that have been developed to serve this purpose have been labeled Content Management Systems (CMS) and have become integrated with document management systems as well as with those dealing with information retrieval. To this we must add that keeping control over processes has become a key element within the current conception of information management. This explains why processes are generally accompanied by workflow systems. The sum of these factors enables us to outline a scenario in which documentary management tools have been incorporating the range of capabilities which are necessary for managing the processes which create, store, treat and present information in the digital environment.

This does not however imply that there exists a total convergence between content management systems and document management systems. On the market we may find content management systems which do not provide the correspondingly desirable documentary capabilities and vice versa. It is also impossible to consider content management and record management to be identical. Yes they are intimately linked, but they are not the same. It is only by studying and evaluating the features and capabilities of available tools that it is possible to determine whether they can provide an adequate solution for a specific problem or context. This issue is further complicated by existing differences between web content management and content management for companies; literature and commentary about these can be found in specialized reports and documents published by companies working in this sector. Moreover, the growing complexity of many organizations' portals, both internal and external, which are supported by content management systems, as well as the digital publishing processes necessary for their production,

makes content management systems develop as mutilayered applications which often reach high levels of complexity.

On the human side, content management tools can only be understood within a collaborative and distributed work environment, for it is only with this kind of entourage that their full potential can be realized. From the reading of the previous paragraphs it can be deduced that, as opposed to traditional passive document treatment approaches , the more up-to-date tools have widened the limits of record treatment, extending all the way from the creation process for its initial stages, and reaching the processes of new publication and personalization for its final stages. This leads to the configuring of a circular or iterative diagram for documentary treatment. To this we must add the necessary reflection over the concept and features of digital documents. The life cycle of digital documents shows significant changes over the traditional life cycle of documents, with the concept of document itself evolving in much the same way. It has become obvious that the way in which informative- documentary activities are carried out must be reformulated in order to face the new digital environment.

2. Content Management Activities and Processes

Content management must be defined from the point of view of its objective and its activities. As opposed to document management, the main focus of which revolves around working with documents, content management is oriented to the management of objects which act as components of virtual documents; this occurs in the context of what we call segmentation. Objects are thus treated through a set of structured processes with the aim of producing digital publications based on the document as a metaphor. This objective of producing digital documents and publications by integrating components can be found in the main references on content management (BOIKO, 2001; BROWNING y LOWNDES, 2001). As with other methods and techniques directly related to the development of the Information and Knowledge Society, theoretical formulations have followed necessity and technological problem solving or have advanced in parallel to these. It is in this sense that in its earliest stages content management has benefited from techniques and processes for digital publication production and editing. It soon became obvious however, that the process of digital publication could not completely cover all the information management processes required for an adequate management of contents.

The most simple digital publishing process defines a process by which an integration framework for a set of objects is designed, its final result being the publication of a digital document. This process, which may reflect the most simple diagram of the functioning of a weblog, becomes increasingly complex if placed in the context of a unit or information system which must create more sophisticated digital information; and which must also serve communities of specialized users or different user communities. As digital information products become increasingly complex so do the technical processes which they support (ROBERTSON, 2003). Collaborative work requires the development of workflows in which it is necessary to assign different roles to the people who are part of them. The objects of which documents are made up may originate from different sources and have a variety of formats; this makes it necessary to control information repositories. Furthermore, the growth of user demand for support services for digital information services, such as information searches or access to documents previously generated, has lead to a growing need for digital archiving tools. Increasingly, content management processes require the formalization of metadata systems which have become necessary not only for management, but for publication and information retrieval also.

This complexity has lead to the development of the two main areas of specialization within content management: Web Content Management (WCM) and Enterprise Content Management (ECM). The first of these (MILLER y DAVID, 2002) deals with the Internet environment and its objectives and methods are focused on the production of digital documents and information for the Internet , particularly for portals and corporative websites. Enterprise content management (GINGELL, 2003) on the other hand is based on the idea of total information management within organizations

through the integration of all the information necessary for the organization to reach its objectives; deriving from publishing systems, ERP systems, record management, data warehouses , etc. Intranets and the internal portals of organizations are currently the most generalized expression of ECM. Labeling of information in XML format, or in specific languages for each context deriving from XML, is a common feature to all content management. To this we might add the increasing need for management of what has been termed digital assets, multimedia documents in digital format which require storage, description, integration and management, with special reference to the legal aspects inherent to these. Another specialized focus is the one that considers tools for the development of digital libraries as content management systems, as is the case in Han's recent paper (2004) comparing Dspace and Greenstone.

Studies on content management tend to agree in pointing out that a minimum requirement for a system of this kind is that it offer a central CMS application supporting publishing, workflow and information repository processes; an information repository; tools for the integration of external information; and templates and models for final products. A CMS is composed of various subsystems that interact with these(BOIKO, 2001):

- Collection: Subsystem which handles the creation and/or acquisition of information. It must provide support for the processes of content creation, workflows, syndication and integration of external sources. In addition, it must provide support to conversion processes between different formats as well as for the incorporation of contents from different sources within specific structures.

- Management: Subsystem in charge of the management and control of information repositories, user groups, and support processes for other subsystems. It handles the defining and controlling of information flows used by other subsystems, as well as the definition of parameters for the functioning of the system.

- Publishing: Subsystem in charge of final production of publications and digital information products in an automatic or semiautomatic manner. It makes use of a model based on templates and must provide personalization options for users as well as the possibility of producing for a variety of platforms and/or customers.

Nakano (2002) points to four subsystems, instead of three, relating to the creation/editing of content, repository, workflow and operations management. Browning and Lowndes (2001) emphasize authorship/creation, workflow, storing/repository and publishing. In any case the dynamics of the products and applications that fall within the scope of content management has made possible the flourishing of various tools following different approaches and which in consequence provide different kinds of solutions. The importance that the choice and implanting of these kinds of tools has for any organization has lead to the carrying out of detailed studies evaluating the features and capabilities of available products. The basic capabilities which these systems must provide may be deduced from these CMS evaluation toolkits. See for example those of Step Two (2004), Gilbane Report (BLUEBILL, 2003), CMS-Watch (2005), or the CMS-Spain guide (CMS-SPAIN, 2004).

Selecting, implanting and making operational a content management tool results from the study and detailed analysis of the organization installing it, of the organization's goals, of the work processes and information resources which it uses, and the users for whom it is intended. A successful content management initiative will therefore require exhaustive planning, preinstallation analysis results, and a rigorous methodology for installation. One study of reference in ECM is Rockley's monograph (2003), which is dedicated to aspects related to analysis, planning, and selection methodology; as well as integration, implanting and making operational. A significant portion of Nakano's text (2002) deals with organization and dynamisation of collaborative work in communities structured through the Internet. ASILOMAR (2003) has published an interesting survey on problems which CMSs cause to their users. As Veen (2004) has pointed out, the success

or failure of a CMS in an organization is not so much a technological issue as one of people and processes.

3. Open software Solutions for Content Management

Open software is a phenomenon which is called to revolutionize business models in the software industry. This article however will not attempt to carry out a detailed analysis of the world of open software, its technological platforms or the typology of existing licenses , as an abundance of information resources dealing with this matter are available elsewhere. For the purposes of this text it will be sufficient to establish the basic criteria that must be fulfilled by the tools we are referring to; these tools must:

- Supply the application source code for the application.
- Be distributed under one of the licenses of reference (see available licenses at Open Source Initiative, http://www.opensource.org
- Be capable of being modified, copied, and distributed freely, in a manner that is respectful of the terms established by the corresponding license.

A brief summary of available content management tools distributed under open software licenses shows the existence of a large number of these offering a variety of capabilities and orientations. In summary, their technical architecture is based on the triad made up of web server, programming language interpreter and database manager. A good example of this diagram is the well known LAMP acronym (Linux, Apache, MySQL, PHP), or its Windows version WAMP. The open content management systems that have become most extended in their use have in fact been PHP (http://www.php.net) and MySQL (http://www.mysql.com), as a majority of solutions are based on these. Examples of all this may be found in OpenSourceCMS (http://www.opensourcecms.com/), and in *CMS Matrix* (http://www.cmsmatrix.org/). The later supplies a very useful and exhaustive comparison matrix for comparing the requirements and features of the range of existing tools. Many of these compare favorably with ownership based high cost solutions and are currently being used in numerous intranets and portals of all kinds of organizations. This has enabled the development of a market and a business model which for the moment appear to be sustainable in both the the short and medium terms. In the last edition of the *LinuxExpo* , Mambo (http://www.mamboserver.com), a CMS under a GPL license obtained the prize for the *Best Open Source Solution* beating the products of large corporations. An analysis of the interest in this kind of solutions on behalf of both customers and added value services providers can be seen in *Holst (2001), Dolye (2003) y Robertson (2004).*

4. A Proposed Typology

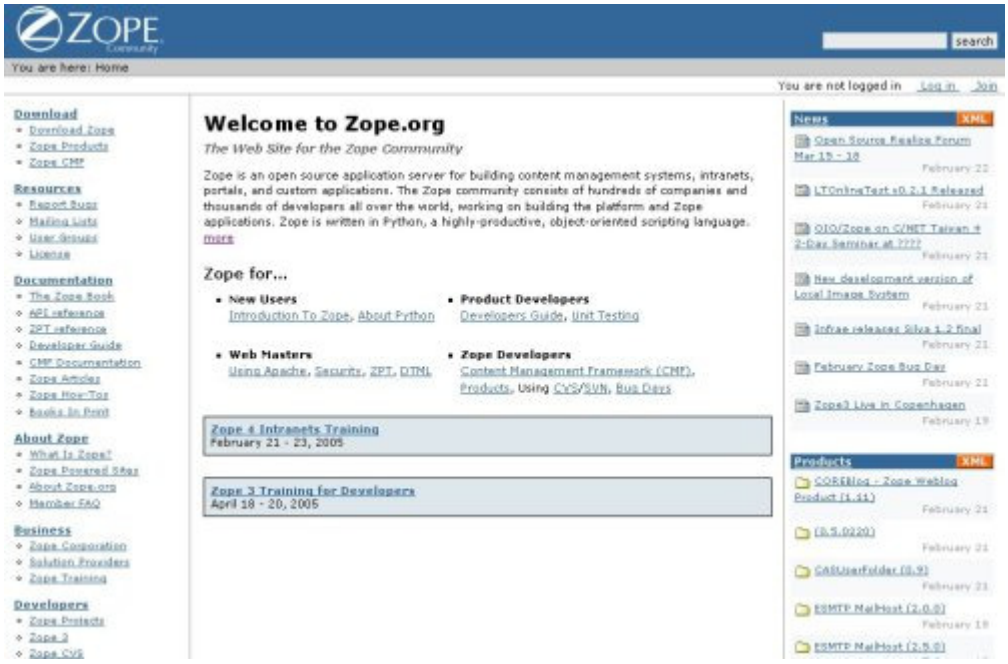
Content management is a rising discipline supplying constantly developing new tools and an increasingly sophisticated theory and methodology. Within the scope of this meta term there exists a diversity of tools and platforms reflecting different orientations, and targeting different objectives and user groups. For the purposes of this article we have decided to create a simplified typology.

Platforms for content management : Solutions offering the necessary platform for developing and implementing applications providing solutions to specific problems. These offer an environment and development tools. The most interesting aspect of these is therefore the possibility of building personalized solutions on a case by case basis. These demand a high level of knowledge of advanced programming tools such as Java or Python. They can be used to build content management and e-commerce solutions or for building tools responding to the features of the types which are explained in the following sections. In many of these, different users and developers have been creating modules making it possible to integrate their own development with

available tools. The features of each platform show a large diversity, for we can find from only the platform, as is the case for Zope, to environments supplying the basic capabilities required for content management, with interfaces and flows specifically designed for this purpose, such as Open CMS. Although this listing doesn't have the pretense of including them all, the most widely used platforms are:

- Zope, http://www.zope.org/
- Typo3, http://www.typo3.org/
- Midgard Project, http://www.midgard.project.org/
- OpenCMS, http://www.opencms.org/
- Apache Lenya, http://lenya.apache.or

Fig 1. Zope



Portals : The creation and upkeep of portals as support and basic tools for web information services is one of the main functions carried out by portal content management systems. Its functionality, administration and control mechanisms are specially designed offer its users a portal containing a variety of contents and services. Ranging from news publishing to document repositories, and including forums, surveys, content syndication, creation of profiles and user groups, personalization of information and its presentation, etc. They follow a modular architecture for they are made up of modules dealing with different functions which are administrated from a centralized interface. For this they make use of the templates mentioned earlier in this article and in which the various modules are distributed. Technically most of these require LAMP/WAMP, even though some tools make use of Python (Plone over Zope) or Perl (such as WebGUI). The most widely used of these are:

- PHP Nuke, http://phpnuke.org/
- Drupal,http://www.drupal.org/

- Mambo, http://www.mamboserver.com/
- Plone (requires Zope), http://www.plone.org/

Fig. 2. Mambo, http://www.mamboserver.com/

Virtual Classroom : Environments offering the necessary capabilities for creating contents for online learning and certain interactive mechanisms such as forums, chats, interactive evaluation, etc. Most of these, rather than being proactive virtual learning environments, are in fact classic classroom models transferred to web environments. This kind of platform has adapted its content management capabilities to the features of digitally formated teaching materials. It is highly advisable that the chosen platform be capable of packaging its contents in SCORM and/or IMS. The most well known platforms are:

- Claroline, http://www.claroline.net/
- Moodle, http:/ /www.moodle.org/

Digital libraries : Digital libraries constitute the paradigm of information services for the end of the 20 th century and beginning of the 21 st . They are organised around the triad formed by users, collections, and value added services, and are developing as a highly specialised area for content management. The development of digital collections, the organisation and creation of access mechanisms or metadata management benefit from the combination of processes of digital publication and the principles of information management. Moreover, they tend to configure collaborative spaces through the access to, and management of, distributed collections. The current state of the main tools varies greatly, as does their focus, for they range from federated record repositories (such as Fedora) to end user tools (such as Greenstone). This kind of tool is more demanding in its requirements than the other kinds discussed in this text.

- Fedora, http://www.fedora.info/
- Dspace, http://www.dspace.org/

- Greenstone, http://www.greenstone.org/

Fig. 3. Greenstone

Digital publications : These are platforms which have been specifically designed taking into account the needs of digital publications, such as newspapers, magazines, etc. Their features and capabilities are therefore essentially concerned with the control of content edition, creation and publishing, rather than with the development of applications or portals. The development of systems dedicated to providing support for sending, reviewing, and publication processes for the publishing of scientific journals will be extremely interesting over the next few years. Although it is not exactly a system for journal publishing we have chosen to include eprints in our listing for its relevance in the academic field and because it offers the possibility of being adapted for publications.

- Cofax, http://www.cofax.org/
- Open Journal Systems, http://www.pkp.ubc.ca/ojs/
- ePrints, http://www.eprints.org/

Environments for colaboration : These would correspond to the classical definition of groupware, tools for working in group. Their aim is to equip specialised user groups with the capabilities needed to work and develop projects jointly. In this kind of environment workflows, users, control points and deliverables are the key contents requiring management. Tools for controlling times and activities are also needed, as well as options for sinchronic and a synchronic communication. We are therefore dealing with an obvious area specialization within content management which is related to project management. One well known and widely diffused collaborative environment is *wik* i. Its main objective being the elaboration of documents for interest sharing communities. A large number of existing wiki platforms have been progressively

increasing their capabilities in such a way that some authors place them in the category of portal tools while others consider the as part of this section:

- eGroupware, http://www.egroupware.org/
- phpCollab, http://www.php-collab.org/
- Wiki (referrence website), http://www.wiki.org/

Blogs : Blogs are the by now well established phenomenon which has given the internet a truly democratic potential. Blogs reflect a very simplified content management system, for they are generally single user and deal with a simple workflow. This has facilitated their expansion amongst a large number of user groups lacking in-depth technical skills. Although the big blog servers provide the user with an already-implemented solution, there exist tools for blogs within de LAMP/WAMP model which can be installed and operated under an open source licence. We must however note a rising trend to increase blog capabilities; these are attempting to give them a set of features similar to those available in portal tools.

- WordPress, http://wordpress.org/

Fig. 4. WordPress

5. Development Perspectives

Content management and the open source software tools which facilitate its implementation are a reality which cannot and should not go unnoticed. They provide solutions for the services that information units must provide in the digital environment. Both general and specialised solutions are available. Moreover, product and support markets for these tools are expanding rapidly. In

order to conclude however, it will be necessary to outline several critical factors which should be taken into account.

- Requirements: Although nowadays it may be possible to set up a service with these tools with only a medium level technical know how, in the near future this technical requirement will increase, this is particularly due to the growing volumes of information, to the complexity of digital documents, and to increasingly demanding security levels.

- Documentation: The professional support for these solutions is a hidden cost which must be measured. Administration and end user documentation in many cases don't currently reach the demanding levels that would be desirable.

- Aims and planning: Against what may be imagined, Internet services in which objectives and correct planning are blurred by the technological mirage are still been developed. It is only through proper analysis and planning that the most adequate choice of tools can be determined: 'method follows function'.

- Semantic Internet: Despite the development of XML and its derived languages, as well as that of information and metadata resources description schemes , content management solutions working directly with these are still few in number. Hardly no support exists as yet for working with RDF or with ontologies. In a digital environment confronted with with the exponential growth of information this is an issue that must be responded to rapidly. The first tools to do this will acquire a more than distinctive competitive advantage.

6. References

ASILOMAR INSTITUTE FOR INFORMATION ARCHITECTURE (2003), *The Problems with CMS* (available in http://aifia.org/pg/the_problems_with_cms.php ; viewed on 12-11-2004).

BLUEBILL ADV. (2003), The Classification & Evaluation of Content Management Systems.En: *The Gilbane Report* , vol. 11, n. 2 (available in http://www.gilbane.com/gilbane_report.pl/86/The_Classification __ Evaluation_of_Content_Management_Systems.html ; viewed 22-1-2005).

BOIKO, B. (2001), *Content Management Bible* . Wiley.

BROWNING, P. y LOWNDES, M. (2001), *JISC TechWatch Report: Content Management Systems* . TechWatch Report TSW 01-02, The Joint Information Systems Committee.

CMS-SPAIN (2004), *Guía CMS-Spain 2004* . (available at http://www.ecm-spain.com/interior.asp?IdItem=1851 ; viewed on 24-11-2004).

CMS-WATCH (2005), *The CMS Report: In-Depth Analysis of Web Content Management Solutions* . (available at http://www.cmswatch.com/TheCMSReport/ ; viewed on 3-2-2005).

GINGELL, D. (2003), *A 15 Minutes Guide to Enterprise Content Management* . Documentum Inc.

GUPTA, V.K., GOVINDARAJAN, S., JOHNSON, T. (2001), Overview of content management approaches and strategies.In: *Electronic Markets* , vol. 11, n. 4, p. 281-287.

DOYLE, B. (2003), Open Source Content Management Redux.En: *The Gilbane Report* , vol. 11, n. 3(available in http://www.gilbane.com/gilbane_report.pl/87/Open_Source_Content_Management _Redux.html ; Viewed on2-2-2005).

ERP Software (2003), *Content Management Tutorial* , (available in http://erptoday.com/CMS/Content-Management-Tutorial.aspx ; viewed on15-01-2005).

GILBANE, F. (2000),Whats content management? In: *The Gilbane Report* , vol. 8, n. 8(disponible en

http://www.gilbane.com/gilbane_report.pl/6/What_is_Content_Management ; consultado 24-1-2005).

HAN, Y. (2004), Digital content management: the search for a content management system. In: *LibraryHi Tech* , vol. 22, n. 4, p. 355-365.

HOLST. S., (2001), Open Source Content Management: A Parallel Universe? In: *The Gilbane Report* , vol. 9, n. 4 (available in http://www.gilbane.com/gilbane_report.pl/1/Open_Source_Content_Management_A_Parallel_Universe.html ; viewed on22-1-2005).

MCKEEVER, S. (2003), Understanding web content management systems: evolution, lifecycle and market. In: *Industrial Management & Data Systems* , vol. 103, n. 9, p. 686-692.

MILLER, B. y DAVID, D. (2002), *Directions in Web Content Management.* Burntsand, Inc.

NAKANO, R. (2002). *Web Content management. A Collaborative Approach* . Prentice Hall.

ROBERTSON, J. (2003), So, what is a content management system?· En: *KM Column* , June 2003 (available in http://www.steptwo.com.au/papers/kmc_what/index.html; viewed on1-2-2005).

ROBERTSON, J. (2004), Open-source content management systems. En: *KM Column* , January 2004, (availble in http://www.steptwo.com.au/papers/kmc_opensource/index.html ; viewed on 12-2-2005).

ROCKLEY, A. (2003), *Managing Enterprise Content. A Unified Content Strategy* . New Riders.

STEP TWO (2204), *Content Management Requirements Toolkit,* Step Two, Sydney.

VEEN, J. (2004), *Why Content Management Fails* . Adaptive Path. (Available in http://www.adaptivepath.com/publications/essays/archives/000315.php ; viewed on 16-10-2004).

WHITE, M. (2002), Selecting a content management system. In: *VINE* ,vol. 32, n. 2, p. 34-39.

7. Links

AIIM Enterprise Content Management Association, http://www.aiim.org/

EDOC Magazine, http://www.edocmagazine.com/

CMS Matrix, http://cmsmatrix.org/

CMS Professionals, http://www.cmprofessionals.org/

CMS Review, http://www.cmsreview.com/

CMS Watch, http://www.cmswatch.com/

CMS Wiki, http://www.cmswiki.com/tiki-index.php

CSM Wire, http://www.cmswire.com/

CSM Spain, http://www.cms-spain.com/

OSCOM, http://www.oscom.org/